

# Python-Based Data Analysis Tool for a 6-DoF Industrial Robot

A Alsabbagh<sup>1</sup>, A N Nasser<sup>1</sup> and G Husi<sup>1</sup>

<sup>1</sup> Department of Mechatronics, University of Debrecen, 4028 Otemeto utca 2.-4., Debrecen, Hungary

[Ammar-alsabbagh-eng@hotmail.com](mailto:Ammar-alsabbagh-eng@hotmail.com)

**Abstract.** A newly developed tool for analysing and visualizing data based on python. The main aim of this tool is saving the data analysis time by automating the data analysis process and creating a Graphical User Interface (GUI) to facilitate the usage of this tool regardless how much they experience programming with Python. This paper's objectives are twofold. First, the paper discusses the characteristics that make the Python language a preferable language for data analysis applications. The paper then describes the architecture of the developed software tool and provides a variety of examples to show the developed tool's advanced features and functionalities. This tool is developed using open source libraries (Pandas, Matplotlib, Tkinter, and SciPy) to manage multi data files at once using one tool, as well as visualizing and storing these data in one file in a clean and organized way.

## 1. Introduction

In R&D, process development department, one of the important steps while building the production line is simulating the assembly process machines in order to define and determine the configuration of the production line and the part managements. For this purpose, Battenberg Robot is used, which is a 6-DoF (Degree of Freedom) Robot that is integrated with 6-Axis Force and Torque sensors (ATI-Sensors) for measuring the required forces and torques of the assembly processes.

The described Robot exports the sensor measurements, the robot end effector position and orientation, and the angles of the robot motor's joints to text files, (one file for each single movement). To manage all those generated txt files, in an automated way, data analysis tool was developed using Python language.

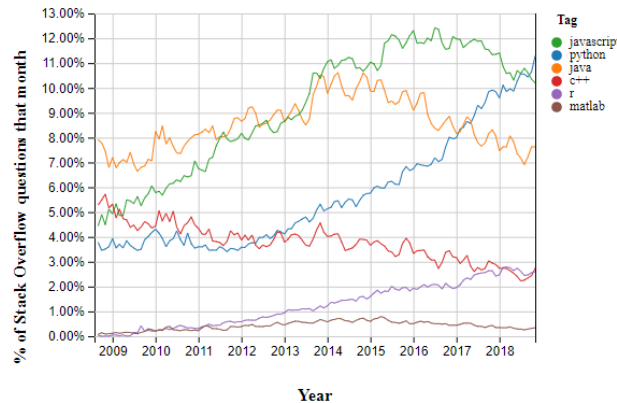
## 2. Why Python over all other languages

It is an easy language to learn, that means developers can build their complete applications even if they are not professional programmers. Moreover, it is becoming progressively used by data scientists, which supports choosing this language for this specific application.

Python is simple because its code is highly readable. Even someone who has no experience with programming can understand most of the code written utilizing Python. One additional fact is, It does not use pointers which means that the chances of program crash are reduced.

Memory management in python is automatic. There is no need to delete any objects, variables etc. in python to free up memory space. Memory de-allocation is done periodically. Python has a wide range of built on functions which tends to simplify most processes. In python variables do not store a fixed data type and can store any data type unlike languages like C++.

Finally, according to the most popular Q&A blog for programmer (Stackoverflow) python's tag trend is becoming dramatically more and more used than other programming languages like Java, JavaScript, C++, R and MATLAB over last 10 years from 2009 to 2018 (see the figure 1).



**Figure 1.** Stack Overflow trends of most popular programming languages over last 10 years

### 3. The data analysis tool functions

#### 3.1. The data analysis tool capabilities

The developed tool is capable of doing all the following functions using only the text files that are generated by the Robot (see the block diagram in Figure 2).

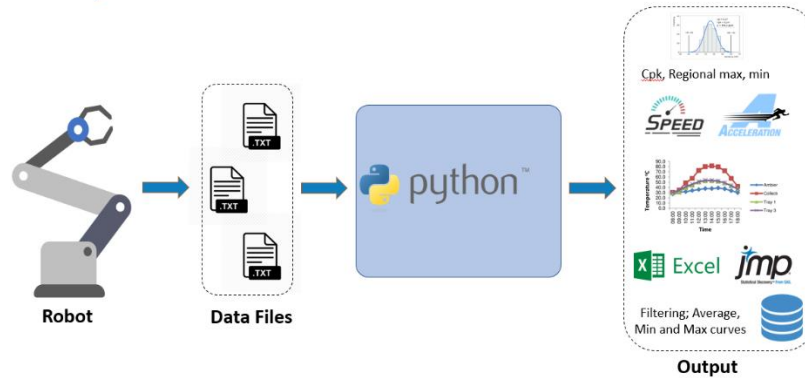
- Calculate the velocity and the acceleration (both linear and rotational) of the robot end-effector
- Visualize the data on two axis coordinates X, Y.
- Calculate and plot the average curve of all selected files, in addition to finding the maximum and minimum curves.
- Applying the Savitzky Golay filter for smoothing the visualized data.
- Visualizing two or more different data sets on the same figure for an easy comparison.
- Zooming in/out in both dynamic and constant ways.
- Save the graph for most of possible image extensions (png, jpg, pdf, etc).
- Changing the axis labels and the figure title.
- Calculating the minimum and maximum in a specific region of the all graphs (regional min, max).
- Calculate the Process Capability Index Cpk, for a specific point of the process.
- Showing the calculations on the same GUI
- Export the row data, filtered data, calculated average, maximum and minimum to sets to an Excel file xlsx and Coma Separated Value file csv

#### 3.2. The tool steps of analyzing the data

The steps of data analysis tool can be seen in figure 3. It starts by opening the text files until the final GUI is open are:

- a. 'Open window' pops up for the user to choose the files that should be analysed.
- b. First user action is required for choosing the file/files that should be managed by the tool
- c. The chosen files are opened one by one, and the data sets are shown in the First GUI, allowing the user to select the required data sets of the experiment.
- d. Second user action is required for selecting the data that is required to be analysed, visualized, saved to the data base.
- e. Organize all the data sets of all the selected files in one data frame (each data set is linked to the file's name).

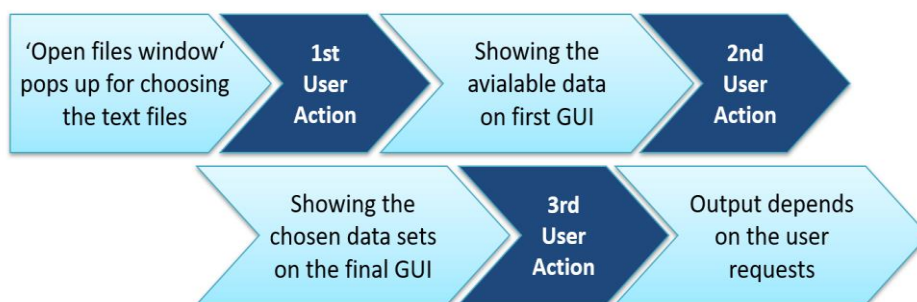
- f. The Final GUI pops Up
- g. When the final GUI is showed, the user can do the data analysis will have all the data analysis option that is provided by the GUI.



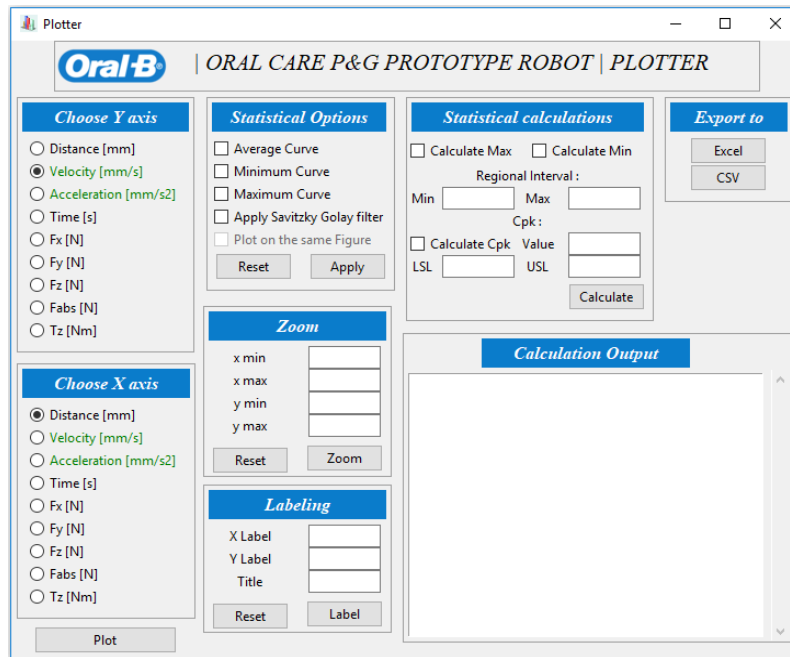
**Figure 2.** Block diagram of the data analysis tool.

When a user runs the program, the “open file window” pops up, allowing the user to select whatever the required files to be analyzed. With this window, the user can select one or more text files at the same time. Then the program opens those text files and checks automatically the available data sets in those files (the data sets are changeable according to the robot program and sensors configuration), then it shows the available data on the first graphical user interface

The first GUI allows the user to select among of the available data. When the user selects the required data to be analyzed, these data will be cleaned, organized and added to one data frame, taking in consideration the change of the text files according to the change of the time and type of assembly test (number of text lines, the size of the data, the type of the robot movement, the location of the data sets header, the number of the text lines at the end of the text file, and etc.). All the transitional velocity and transitional acceleration of the robot end-effector are calculated in this stage, or the rotational velocity and rotational acceleration of the robot end-effector are calculated in this stage according to the type of movement. Moreover, the set velocity and acceleration of the robot are read by the program to compare them with the actual velocity and acceleration set by the robot program.



**Figure 3.** Data analysis tool steps.



**Figure 4.** Final Graphical user interface for visualizing the data

Figure 4 shows the final graphical user interface for analyzing the data. This GUI contains all the tool Functions. Both sections “Choose Y axis” and “Choose X axis” are changeable according the selected data set in the Previous GUI (First GUI), while all the other section on this GUI (Statistical Options, Zooming, Labeling, and etc.) are graphically constants. The user can choose the output according to documentation requirements. The Visualization will be on a separate window, then the users can have as many figures as they demand.

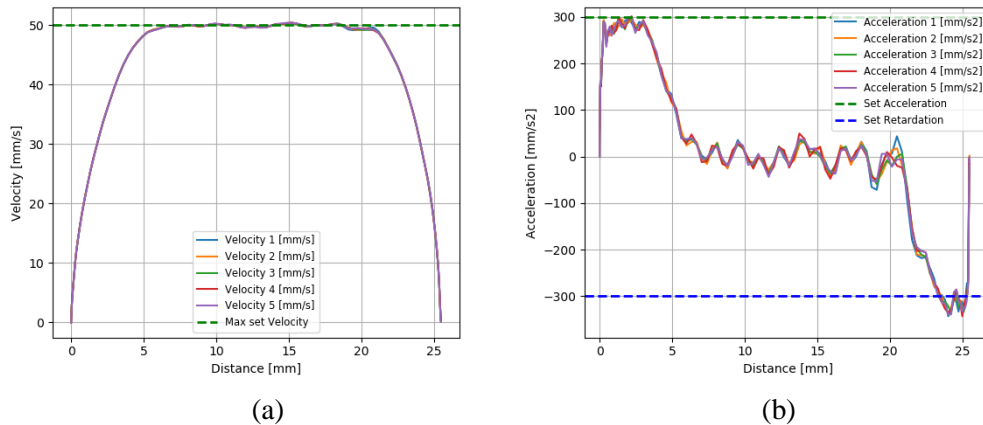
The axes selection sections enable the user to choose the X axis and Y axis for visualizing the data. The data in Black is the row data that come directly from the robot text files, while the data in green is the calculated data. The tool can calculate all the following data according to the robot movement type: transitional velocity, transitional acceleration, angular velocity, and angular acceleration.

$$v = \frac{x_2 - x_1}{t_2 - t_1} \quad (1)$$

$$a = \frac{v_2 - v_1}{t_2 - t_1} = \frac{x_3 - 2x_2 + x_1}{(t_2 - t_1)^2} \quad (2)$$

$$\omega = \frac{\theta_2 - \theta_1}{t_2 - t_1} \quad (3)$$

$$\alpha = \frac{\omega_2 - \omega_1}{t_2 - t_1} = \frac{\theta_3 - 2\theta_2 + \theta_1}{(t_2 - t_1)^2} \quad (4)$$



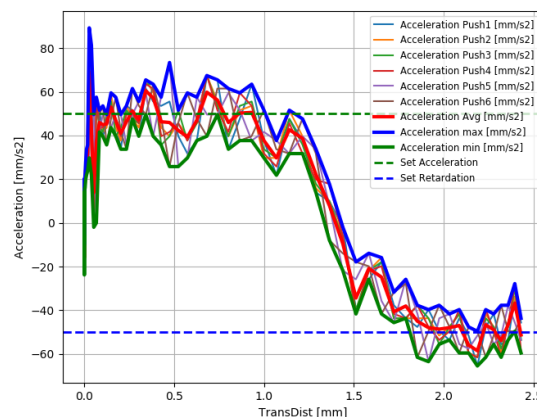
**Figure 5.** The Robot end-effector velocity and acceleration

Figures 5 (a) and (b) show the comparison between the actual end-effector velocity with the set speed (a), and the same for the acceleration and the retardation (b). This feature is very significant for the assembly test, as setting the velocity of the robot should be exactly the same of the velocity of the assembly machine of the production line. With this feature the user can check the speed validity of the robot. And in case the robot is not reaching the desired velocity, an action should be taken, increasing the acceleration for instance.

#### 4. Statistical Options

The “Statistical Options” section enables the user to have variety of statistical options for the visualization as well as for defining some of the most important curves. All can be done by selecting the required option among these section options. With this section the user can do all the following:

1. Calculate the average curve of all the selected text files, add it to the data frame and plotting it on the figure.
2. Define the minimum curve among all the curves that come from the selected files, add it to the data frame, and potting it on the figure.
3. Define the maximum curve among all the curves that come from the selected files, add it to the data frame, and potting it on the figure. Figure 6. Show an example of calculating the average of acceleration signals that are coming from different six test files, in addition to the finding the maximum/minimum of those six files, then visualizing them.

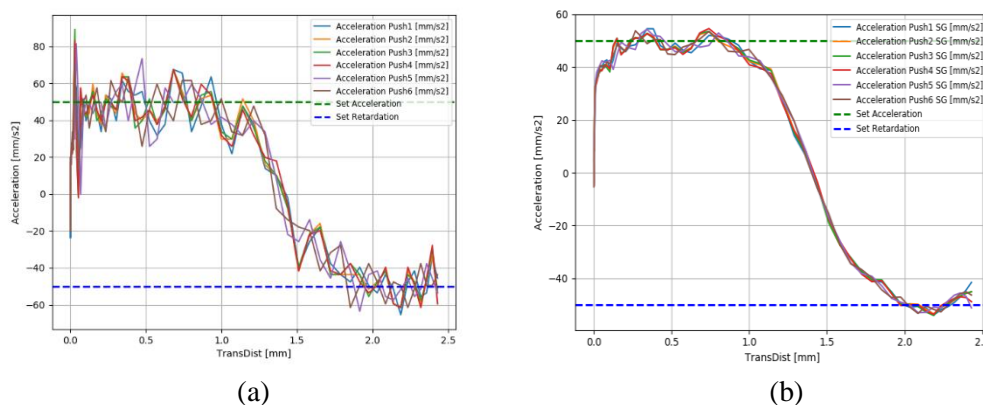


**Figure 6.** the visualization of Average, minimum and maximum on the top of the raw data.

4. Smoothing the raw data using Savitzky Golay filter, add the filtered data to the data frame, and plotting it on the figure.

A Savitzky–Golay filter is a digital filter that can be applied to a set of digital data points to smooth the data, i.e. to increase data accuracy without distorting the tendency of the signal. This is achieved by fitting successive sub-sets of adjacent data points with a low-grade polynomial using the linear least squares method. When the data points are equally spaced, an analytical solution for the least square equations can be found in the form of a single set of "convolution coefficients" that can be applied to all data sub-sets to provide estimates of the smoothed signal (or smoothed signal derivatives) at the center point of each sub-set. The method was popularized by Abraham Savitzky and Marcel J. E. Golay, who published tables of convolution coefficients for various polynomials and sub-set sizes in 1964 [1][2], based on established mathematical procedures. Some tables errors have been fixed [3][4]. The method for handling 2-and 3-dimensional data has been extended [5].

figures 7 illustrate the acceleration of six different text files before (a) and after applying Savitzky Golay filter (b). The figure after applying SG filter, can show that the singles are smoothed and more understandable. That can be helpful for understanding the complete trend of the signals.



**Figure 7.** the acceleration before and after applying Savitzky Golay filter

## 5. Statistical Calculation

A process capability index is a numerical summary that compares the behavior of a product or process characteristic to engineering specifications. These measures are also often called capability of performance indices or ratios; however, the generic term is capability index. A capability index relates the voice of the customer (specification limits) to the voice of the process. A large value of the index indicates that the current process is capable of producing similar parts, that will meet or exceed the customer's requirements. This term is widely used in the process because it reduces complex information about the process to a single number. There are various applications for the capability index, though the use of the indices is driven mostly by monitoring requirements specified by customers. Many customers ask their suppliers to record capability indices for all special product characteristics on a regular basis. These indices are used to communicate how well the process has performed. It is assumed that these indices also indicated expected future performance. Suppliers may also use Cpk for different characteristics to establish priorities for improvement activities. Similarly, the effect to process change can be assessed by comparing capability indices calculated before and after the change.

If the upper and lower specification limits of the process are USL and LSL, the target process mean is T, the estimated mean of the process is  $\hat{\mu}$  and the estimated variability of the process (expressed as a standard deviation) is  $\hat{\sigma}$ , then commonly accepted process capability indices include:

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}} \quad (5)$$

Estimates what the process is capable of producing if the process mean were to be centered between the specification limits. Assumes process output is approximately normally distributed.

$$\hat{C}_{p,lower} = \frac{\hat{\mu} - LSL}{3\hat{\sigma}} \quad (6)$$

Estimates process capability for specifications that consist of a lower limit only (for example, strength). Assumes process output is approximately normally distributed.

$$\hat{C}_{p,upper} = \frac{USL - \hat{\mu}}{3\hat{\sigma}} \quad (7)$$

Estimates process capability for specifications that consist of an upper limit only (for example, concentration). Assumes process output is approximately normally distributed.

$$\hat{C}_{pk} = \min \left[ \frac{USL - \hat{\mu}}{3\hat{\sigma}}, \frac{\hat{\mu} - LSL}{3\hat{\sigma}} \right] \quad (8)$$

Estimates what the process is capable of producing, considering that the process mean may not be centered between the specification limits. (If the process mean is not centered,  $\hat{C}_p$  overestimates process capability.)  $\hat{C}_{pk} < 0$  if the process mean falls outside of the specification limits. Assumes process output is approximately normally distributed.

$$\hat{C}_{pm} = \frac{\hat{C}_p}{\sqrt{1 + \left(\frac{\hat{\mu} - T}{\hat{\sigma}}\right)^2}} \quad (9)$$

Estimates process capability around a target, T.  $\hat{C}_{pm}$  is always greater than zero. Assumes process output is approximately normally distributed  $\hat{C}_{pm}$  is also known as the Taguchi capability index [8].

$$\hat{C}_{pm} = \frac{\hat{C}_p}{\sqrt{1 + \left(\frac{\hat{\mu} - T}{\hat{\sigma}}\right)^2}} \quad (10)$$

Estimates process capability around a target, T, and accounts for an off-center process mean. Assumes process output is approximately normally distributed.

Process capability indices are constructed to express more desirable capability with increasingly higher values. Values near or below zero indicate processes operating off target ( $\hat{\mu}$  far from T) or with high variation.

That is why calculating Cpk is a significant function in this GUI, especially for process development team. The data analysis tool is capable of calculating the Cpk for any point of the selected signal, and actually it is used for the most critical point during the executed tests. It should be mentioned, that in the assembly tests, few tests are executed for the same parts (less than a hundred). Therefore, it is mentioned on the output section of the data tool that the Cpk value may not give a conclusion of the basic distribution. Nevertheless, the Cpk feature is still giving a pre-description of the process before it goes



for the production line. The Cpk value and all the other critical values will be showed on the “Calculation output” section.

Finally, one of the most useful features is exist in the “Export section”. In this section there are two buttons for saving the data in two different file types. The First one is xlsx extension which is the extension of the Excel file, and the other extension is CSV Comma separated value that can be operated using different software like JMP, Excel etc.

With this feature the user would be able to save the selected data sets of all the files, the clean data, the calculated data (velocity, and acceleration), the filtered data, the maximum curve, minimum curve, and the average curve in an organized and accessible way, and this data will be linked to their corresponded files. The importance of this feature come from allowing the user storage the data in one SharePoint which is the ground of the database. Moreover, the user would be able to do more advanced data analysis which is not exist in the data analysis tool by using JMP program for example.

## 6. Conclusion

By using this data analysis tool, the time efficiency increased dramatically; more than 90% of data analysis consumed time is saved using the developed graphical user interface. Moreover, multiple tests on this tool were done, asking people who were never introduced to this tool before to do around fifteen tasks. All of them were able to finish all the fifteen tasks in less than 10 minutes, which shows how much this GUI is user-friendly, easy to be used, and sufficient for manipulating the data, moreover the same fifteen tasks had required more than three hours to accomplish them by the traditional way (copy data to Excel or JMP then analyze the data and visualize it). Finally, it should be mentioned that this tool was developed as a fast prototype for test and demo tool. Which means, more development is required, and all the cybersecurity issues should be taken into consideration for getting a safe and reliable version.

## Acknowledgments

The publication is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund.

## References

- [1] Whittaker E T and Robinson G 1940 Calculus of observations: *Treatise on Numerical Mathematics 3rd Edition* Blackie and Sons Limited, London.
- [2] Guest P G, and Guest P G 2012 Numerical methods of curve fitting *Cambridge University Press*. pp. 147
- [3] Savitzky A, and Golay M J 1964 Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, **36(8)** 1627-1639.
- [4] Pochinok N B. 2008 Contemporary Problems of Social Work. *Bulletin of Chelyabinsk State University*, **(4)** 88-97.
- [5] Steinier J, Termonia Y, and Deltour J. 1972 Smoothing and differentiation of data by simplified least square procedure. *Analytical chemistry* **44(11)** 1906-1909.
- [6] Larive C K, and Sweedler J V 2013 Celebrating the 75th anniversary of the acs division of *analytical chemistry*: A special collection of the most highly cited analytical chemistry papers published between 1938 and 2012
- [7] Alsabbagh A, SZEMES P T, & Baki A 2018. Brushless DC Motor Modeling Using Bond Graph Method and Control using LabVIEW. (*RIim*) *Recent Innovation in Mechatronics*, **5(1)**
- [8] Boyles R A 1991 The Taguchi capability index. *Journal of Quality Technology* **23(1)** 17-26